Exploring the Effects of Automated Pronunciation Evaluation on L2 Students in Thailand

Simon Moxon
Walailak University, Thailand

Abstract

A significant barrier to effective communication in a second language is the awareness and accurate reproduction of phonetic sounds absent in the mother tongue. This study investigated whether the automated evaluation of phonetic accuracy using speech recognition technology could improve the pronunciation skills of 105 (88 female, 17 male) Thai undergraduate students studying English in Thailand. A pre-test, post-test design was employed using treatment and control sample groups, reversed over two six-week periods. Treatment group students were given access to an online platform on which they could record and submit their speech for automated evaluation and feedback via SpeechAce, a speech recognition interface designed to evaluate pronunciation and fluency. Independent samples t-test analysis of the results showed statistically significant improvement in pronunciation accuracy of students in the treatment group when compared to those in the control group (t (89) = 2.086, p = .040, 95% CI [.083, 3.423]), (t (89) = -4.692, p < .001, 95% CI [-5.157, -2.089]). Pearson's correlation analysis indicated a weak to moderate, but statistically significant correlation between frequency of practise and pronunciation test score (r =.508, p < .001), (r = .384, p = .021). The study has limitations as the sample group was predominantly female, and time constraints limited students' use of the software. Future studies should investigate possible gender differences and experiment with different forms of visual feedback.

Keywords: phonetics, pronunciation, speech recognition, SpeechAce

Although considered the most important of the four language skills for second language (L2) students, impractical student to teacher ratios and a greater emphasis on teaching grammar means speaking is often neglected in the classroom (Gerald, 2000; Leong & Ahmadi, 2017). To effectively correct pronunciation errors, the student must receive corrective feedback at the time of the error (Huang & Jia, 2016), yet this can often prove to be impractical in the classroom unless the lesson's focus is pronunciation and class sizes are manageable. Assuming that receiving feedback in an environment where students felt secure and less exposed might alleviate their anxiety, the present study investigated the influence on students' pronunciation skills after using online speech recognition software. The software is capable of analysing phonetic accuracy and fluency and delivering instant evaluation and feedback. The study sought to answer the following research questions:

1. Can the use of speech recognition software positively influence students' English pronunciation?
2. Is there a correlation between the frequency of use of the software and improved pronunciation test scores?

During this research, the author taught speaking, listening, and phonetics in two-hour sessions to class sizes in excess of 40 students, making one-on-one tutoring impractical. It was observed that the voicing of word ending sounds, such as the /saIz/ of "exercise", and pronunciation of specific phonemes, such as the voiceless dental fricative "th"/θ/ as in "thin", were problematic for many of the students.

English major students at Walailak University have a broad range of English-speaking ability and experience in communicating with native English speakers. Some students studied in English programmes during compulsory education and had regular contact with native English speakers from a relatively early age. Other students studied English in traditional programmes with Thai staff or other non-native English speakers. However, the students with more accurate English speaking and pronunciation skills and overall speaking confidence were not necessarily those who studied in English programmes or with native English speakers. Possible reasons for this may be the insufficient practice of pronunciation or a greater fear of making a mistake in front of a native speaker. Research has shown that a significant factor in students' speaking ability is their self-confidence (Aiello & Mongibello, 2019; Ayulistya, 2016; Leong & Ahmadi, 2017). Limited pronunciation skills can often lead to communication breakdown and raised anxiety (Dewaele, 2007; Dewaele, Furnham, & Petrides, 2008) and a loss of self-confidence (Donovan & Macintyre, 2004; Gilakjani, 2012). Fear of making a mistake or mispronouncing a word in the presence of their peers can lead to anxiety that prevents the student from speaking. Conversely, the teacher may refrain from giving correction for fear of hurting the student's feelings (Huang & Jia, 2016).

Considerable research has attempted to define the indicators of fluent speech (De Jong, 2018; Ejzenberg, 2000; Lennon, 1990; Yang, 2014a). However, based on their survey of 84 L2 teachers teaching English as a Foreign Language (EFL), Tavakoli and Hunter (2018) state that accurate pronunciation, including accent and intonation, is commonly rated as a critical indicator of what Lennon describes as fluency in its broad sense (Lennon, 1990). With increasing demands on students to pass standardised tests such as the International English Language Testing System (IELTS), where pronunciation is a key descriptor of speaking ability (International English Language Testing System [IELTS], n.d.), the students' need to practice and receive evaluation and correction of their L2 speaking is significant. Despite this, L2 students often have limited time inside and outside the classroom in which to practise speaking

with a native speaker (Leong & Ahmadi, 2017) or are taught by non-native speakers who may be lacking in the required L2 teaching skills or grasp of English pronunciation (Kanoksilpatham, 2007).

Literature Review

Pronunciation Barriers

Incorrect pronunciation can often be a barrier to social interaction and lead to misunderstanding (Aiello & Mongibello, 2019; Ayulistya, 2016; Fraser, 2000; Leong & Ahmadi, 2017), which in turn may cause anxiety and loss of confidence in the speaker (Dewaele, 2007; Dewaele et al., 2008; Gilakjani, 2012). Mother tongue influence, such as missing sounds, intonation, and tonal use, are often attributed to causing pronunciation error (Gilakjani, 2011; Jahandar, Khodabandehlou, Seyedi, & Abadi, 2012; Lai, Tsai, & Yu, 2009; Latha & Ramesh, 2012). The listener may also interpret such influences on pronunciation as disfluency in the target language (Brumfit, 1984; De Jong, 2018; Derwing & Munro, 2005; Ejzenberg, 2000; Latha & Ramesh, 2012; Lennon, 1990; Richards, Platt, & Weber, 1985; Yang, 2014b).

Another problem that arises in terms of pronunciation accuracy and perceived accuracy from the listener's perspective (Christiansen, 2011) is that the English language is no longer a single standard. Globalisation has led to many English language strains, with different accents producing different vowel sounds (Hariri, 2012). According to Gilakjani (2012), learners require significant exposure to the target language to be proficient enough to speak. Still, lack of contact with native speakers and exposure to the target language often hinders their progress. Khamkhien (2010) found that problems with Thai students' pronunciation of English words increased with the number of syllables, which reflects this viewpoint. Following his assessment of pronunciation of multi-syllabic words by 90 Thai students, he suggests that the issue derives from word stress and that teaching attention should focus on this area of phonics.

Although exposure to the native language is a significant factor in accurate pronunciation (Kenworthy, 1987), lack of contact and communicative opportunities with a native speaker is commonplace in some Thai provinces. This issue is more noticeable in rural areas and areas less frequented by western tourists, where the student to teacher ratio can render one-to-one communication impractical or impossible (Ngamkaiwan, 2018). Conversely, in cases where students have contact with nationals from different English-speaking countries, watch English movies, or listen to English pop music, it is not uncommon to detect the influence of such exposure on their pronunciation and vocabulary. While the listener may perceive different accents acquired by the L2 students as disfluency or poor pronunciation, Hariri (2012) argues that pronunciation accuracy is not native-like pronunciation but a measure of speech clarity. Her study of related literature looked specifically into the role of gender on pronunciation accuracy. While her findings regarding gender differences in accuracy of pronunciation support other research findings (Hincks, 2003; Jahandar et al., 2012; Khamkhien, 2010), she concluded that teaching style did not influence the gender difference.

Pronunciation Pedagogy

Fraser (2000) argues that even in the presence of other errors, such as grammar, successful discourse can be achieved through good pronunciation. However, research has shown that the teaching of pronunciation in the classroom is often overlooked or allocated the least amount of time and attention (Derwing et al., 2012; Derwing & Munro, 2005; Fraser, 2000; Gilakjani, 2011; Hariri, 2012), or completely absent from curricula after the introductory stage (Gilakjani, 2011). When faced with teaching pronunciation, a lack of professional development,

knowledge in incorporating pronunciation instruction into the classroom, and reduced opportunity to implement pronunciation pedagogy based on research findings, have led to teachers depending more on textbooks and instinct than evidence-based guidance (Derwing & Munro, 2005; Derwing et al., 2012). L2 English, taught by inexperienced non-native speakers, is also a factor in the fossilisation of incorrect pronunciation in students' dialect (Kanoksilpatham, 2007). Derwing et al. (2012) also found notable inequalities in task and teaching between textbooks within a series and a lack of clarity in the explanation for students and teachers. They called for greater integration of pronunciation in the various textbook tasks.

In his research, Gilakjani (2011) concluded that teachers should encourage students to practise their English-speaking skills beyond the classroom's confines. However, while this is good in practice, students are often unable to make appropriate use of the English they learn in class beyond the classroom's boundaries (Gumbaridze, 2013). In Thailand, particularly for students in remote and rural areas, the concept is not so readily and effectively put into practice as opportunities to converse with native English speakers may be scarce and, in the case of practising pronunciation, void of adequate corrective guidance. In the absence of proper correction, speakers may be unaware of their errors, and hence, mispronunciations become fixed in their regular discourse (Hincks, 2003; Swain & Lapkin, 1995). Hincks (2003), who used an application called "Talk to Me (English)", observed that an improvement in pronunciation accuracy was only noticeable in those students deemed to have poor pronunciation attributed to their mother tongue accent.

Correcting mispronunciation and providing remedial help to students has been debated considerably in the research literature. While teachers and students may agree that receiving corrective feedback is necessary (Huang & Jia, 2016), the type of feedback and the timing of its giving can be problematic and create barriers to accurate pronunciation (Heift & Schulze, 2007; Huang & Jia, 2016). According to Gumbaridze (2013), the schedule for and method of giving corrective feedback is a controversial topic regarding teaching methodology as it may be inappropriate, intrusive, or result in demotivating the student. In terms of correcting mispronunciation, Hincks (2003) states that the time required to evaluate pronunciation, and inconsistency between raters, are critical factors in error correction. She also notes that uncertainty on the appropriate form of corrective feedback creates a barrier for teachers in the classroom. Hincks states that fear of receiving corrective feedback in front of their peers may prevent students from speaking in the classroom, an argument supported in other research (Gumbaridze, 2013; Huang & Jia, 2016; Terrell, 1997).

Pronunciation Software

With the wide use of mobile technology in the classroom, students have many tools to practice and assess their pronunciation. Considerable research exists in the area of automatic speech recognition (ASR) software for language learning, and findings have supported some correlation between its use for phonetic training and improved pronunciation (Ayulistya, 2016; Haggag, 2018; Hincks, 2003; Lai et al., 2009; Olson, 2014). However, the research literature also suggests a disconnect between research findings and classroom application (Derwing et al., 2012; Derwing & Munro, 2005; Olson, 2014).

ASR offers several advantages over human evaluation, such as instant feedback (Haggag, 2018), reduced demand on teacher time, consistency of rater, and quantitative feedback on accuracy to a very detailed level (Hincks, 2003). As a tool for students to self-assess their accuracy (Lord, 2005) and gain valuable practice time beyond class hours, ASR has much to offer the L2 student in terms of guidance (Hincks, 2003) and accurate identification of

mispronounced segmentals (Srikanth et al., 2012). However, Kim (2006) expressed concern regarding the reliability of scores generated by a computer algorithm. In cases where the ASR software uses the Hidden Markov Model (HMM), the software compares the sounds received with phoneme sounds stored in a database to find a probabilistic match (Hincks, 2003). While this may have some potential for producing false positives when evaluating pronunciation accuracy, most systems can accurately determine the percentage of deviation between the received and stored sounds and calculate an accuracy rating (Hincks, 2003).

Opinion on the use of ASR as a tool for practising pronunciation is mixed. According to Haggag (2018), ASR promotes autonomous and collaborative learning while offering automatic evaluation and feedback. In his study of 23 trainee English language teachers, he also found that user satisfaction and appreciation in using the software were high. However, Setter and Jenkins (2005) argue that more advanced systems, such as Praat, are intended for researchers and are too complicated for teachers and students to understand. Referring to Praat, Brett (2004) concluded that, as the software was not intended for this form of use by pronunciation learners, the interface is not suitable and needs redesigning. Olson (2014), however, rejects these claims and believes ASR systems, such as Praat, could be successfully used in the L2 classroom. In his research, he found that students reported the visual representation of the phoneme sounds, which the software generates automatically, helped many compare the difference between the pronunciation of non-native and native speakers, while only a few experienced problems using the software. Echoing the benefits regarding the visual representation of sounds, Hincks (2003) found that displaying contours to represent pitch changes was attributed to improved intonation. The software used by Hincks, "Talk to Me (English)", can generate a waveform representation of the sound, similar to Praat, with visual guidance regarding mouth and tongue positions. However, Hincks raises a significant concern, pointing out that a focus on pronunciation using ASR software can lead to students deliberately slowing their speech to emphasise on clarity, consequently having a negative impact on their speaking fluency.

The literature offers much debate regarding the suitability of software, such as Praat, in autonomous learning, but agreement on the effectiveness of waveform representation of sounds generated by such systems diverges. Conversely, there is little in the literature to show that waveform representation alone, i.e., void of any corrective guidance, is any more informative to the student than a numerical rating for each phoneme sound.

Method

Voice Recognition Technology

Prior to this investigation, students were encouraged to use the built-in voice recognition features of their laptops and other mobile devices, such as the voice-to-text and voice command accessibility features in Microsoft Windows, or translation applications such as Google Translate. While these give some form of visual feedback on pronunciation accuracy, that is, the text and speech are the same, they offer no corrective feedback when the text and speech differ. It was also found that the technology could sometimes assume the intended word, possibly based on rhythmic patterns, or assumed context, even when mispronounced, thus giving false feedback to the student. For example, using some translation applications with built-in ASR, a phrase such as "We wish you a Merry Christmas" could be badly mispronounced, yet the ASR could often assume and display the correct version of the text on the screen if the rhythmic aspect of the utterance were accurate enough.

A web-based platform was chosen as it would enable easy access for the typical devices currently used by the students. It would also allow students to access the system remotely while allowing the teacher to update content and monitor usage without manually updating student devices. The SpeechAce application programming interface (API) was selected as it offered the ability to interface with the existing web-based Learning Management System (LMS) used for the course. It also gave feedback on each attempt at sentence, word, syllable, and phoneme levels while enabling the recording of each feedback response in the LMS database. SpeechAce also requires the submittal of a transcript to compare against each submitted sound, which reduced the potential for encountering the false positives observed in sound only ASR applications. The website communicated with the programming interface provided by SpeechAce using a secure socket layer (SSL) (Figure 1). Ten pronunciation exercises were programmed into the LMS each week for students to attempt. In each exercise, the LMS presented students with a short sentence on their screen and recorded their voice while saying the sentence. Students could review and re-record their attempt before submitting for analysis by SpeechAce. For each pronunciation exercise attempt, the SpeechAce interface returned detailed feedback in the form of a structured array packet. After saving it to a database, the LMS parsed each packet for the user to see on their screen (Figure 2). The system highlighted evaluation scores below 70% in graduated shades of red to facilitate the students' interpretation of the feedback and draw their attention to problematic pronunciation.

The LMS was programmed to offer two forms of exercise; assessment exercise, which allowed each student to submit only a single attempt at each sentence, and practise exercise, which allowed unlimited attempts. All students had access to the assessment exercises, but only students in the treatment group had access to the practice exercises.

Figure 1

Process Flow of Voice Recognition Technology

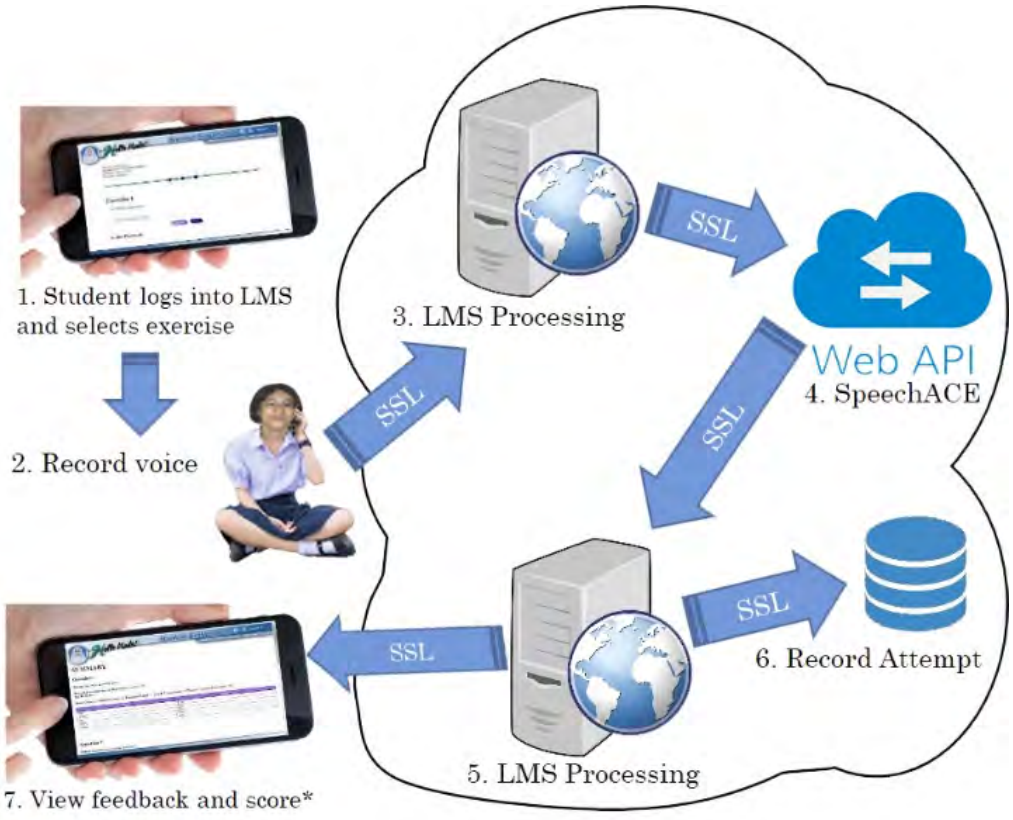


Figure 2
SpeechAce Feedback, Parsed for the User

Evaluation Summary	
Overall score (Average of all phonemes score): 85%	
Word count 6	Syllable count 12
Phone count 28	Fidelity CORRECT

Word	Score
That	61%
is	96%
a	97%
truly	94%
fantastic	82%
activity	85%

Syllable	Score
that	61%
is	96%
a	97%
tru	100%
ly	87%
fan	87%
ta	54%
stic	92%
ac	80%
ti	68%
vi	98%
ty	93%

Phoneme	Score
dh	88%
ae	44.5%
t	49.61%
ih	93.35%
z	98.5%
ey	97.17%
t	99.56%
r	100%
uw	99%
l	74.14%
iy	99.69%
f	97.39%
ae	69.67%
n	94.3%
t	94.4%
ae	12.78%
s	98.5%
t	93%
ih	78.75%
k	96.33%
ae	61%
k	99.33%
t	99%
ih	36.17%
v	100%
ah	97%
t	87.5%
iy	99.38%

Context and Participants

This study was conducted in an English language programme within the Faculty of Liberal Arts at Walailak University, Nakhon Si Thammarat, Thailand. The faculty has approximately 750 students studying in English, Chinese, Thai, and ASEAN programmes at bachelor's degree level. Within the context of this study, the population consisted of first-year English major students aged between 18 and 19 (n = 105, 88 female, 17 male) and studying in the English programme.

Sample

Permission was obtained from the University to complete the study before inviting students to participate. The risks to students participating in this study were no different from attending

class or routinely browsing the Internet. The entire population, 105 students, were invited to participate in the research and advised that their participation was voluntary and that they could opt-out at any point during the 12 weeks. They were also informed that any scores collected during the study would not impact on their course grades. For generalisation purposes, a minimum sample size of 83 was calculated using the formula defined by Krejcie and Morgan (1970), with a confidence level of 95% and a 5% margin of error. Students were randomly assigned to one of two groups, Group A and Group B, and instructed on using the website and SpeechAce interface.

Data Collection and Preparation

The research instrument compares the means of pre-test and post-test scores using t-tests (Hincks, 2003). Data were collected during two separate phases of the study and compared using t-test analysis. First, all students were invited to attempt an assessment exercise, which would form their pre-test scores. Students in Group A (treatment group) were then allowed to complete the practise exercises at their leisure for six weeks while Group B (control group) had no access to the system. At the end of the first six weeks, all students were invited to attempt the second assessment exercise, which would form their post-test scores for the first phase and their pre-test scores for the second phase. The groups were then reversed (Group B now the treatment group), and the study ran for a further six weeks before all students were invited to attempt the third and final assessment exercise, which would form their final post-test scores.

All data were screened for erroneous values and to exclude data from students who did not complete all speaking tasks for each of the three assessment exercises. Fourteen students failed to fully complete all three assessment exercises, leaving data from 91 students usable for further analysis. The remaining data were then screened for violations of the assumptions of the t-test and Pearson Correlation analysis methods. Shapiro-Wilk tests showed no significant departure from normality for Test 1 ($W(91) = .982$, $p = .225$), Test 2 ($W(91) = .990$, $p = .692$), Test 3 ($W(91) = .980$, $p = .178$), and the number of practise attempts for Group 1 ($W(47) = .968$, $p = .221$), and Group 2 ($W(44) = .967$, $p = .230$). No outliers were detected.

Results

Independent samples t-test analysis showed that scores for Test 1 were not significantly different between the groups, (t (89) = 1.396, $p = .166$, 95% CI [-.684, 3.917]), suggesting no significant difference in the level of pronunciation accuracy between the two groups before treatment. The independent samples t-test was calculated to compare differences in pre-test and post-test scores between the two groups. The t-test was significant for the first round of pre-test and post-test, (t (89) = 2.086, $p = .040$, 95% CI [.083, 3.423]) and also the second round of pre-test and post-test, (t (89) = -4.692, $p < .001$, 95% CI [-5.157, -2.089]). An additional t-test analysis showed that the difference between Test 1 and Test 3 scores was also significantly different between the two groups, (t (89) = -2.055, $p = .043$, 95% CI [-3.682, -0.570]). Pearson's correlation coefficient analysis was used to measure the correlation between the students' frequency of practice and the difference between their pre-test and post-test scores. The analysis was statistically significant for Group 1, (r =.508, $p < .001$), and Group 2 (r = .384, $p = .021$).

Discussion

This research set out to address two questions regarding ASR software use on students' pronunciation accuracy. Addressing research question one, asking whether the use of speech recognition software can positively influence students' English pronunciation, the results of this study are encouraging and suggest the use of the speech recognition software did have a positive influence on students' pronunciation test scores, which supports the findings in the research literature (Ayulistya, 2016; Haggag, 2018; Hincks, 2003; Lai et al., 2009; Olson, 2014; Srikanth et al., 2012). In both rounds of testing, the treatment group displayed a statistically significant improvement in pronunciation accuracy scores when compared against the control group. Comparison between Test 1 and Test 3 scores showed that, overall, both groups attained a level of improvement in their pronunciation accuracy with the improvement being more noticeable in the scores for group 2.

Research question two explored the possible correlation between the frequency of use of the software and improved pronunciation test scores. Although the correlation between practise attempts and the difference in pre-test and post-test scores was significant, the strength of the correlation (r_{xy}^2 = .258 for Group 1 and r_{xy}^2 = .147 for Group 2) was considered weak to moderate (Salkind, 2012). It is proposed that repeated practise of the exact pronunciation phrase was not overly helpful in improving pronunciation accuracy because students may have been unable to make appropriate use of the feedback to correct their mispronunciation. This viewpoint reflects the research literature (Brett, 2004; Setter & Jenkins, 2005). In the present study, the feedback was given as percentages against each word, syllable, and phoneme and highlighted in shades of red to reflect degrees of inaccurate pronunciation. While the extent to which students examined the feedback is not apparent, limitations in the students' phonetical knowledge could have made it too difficult for many of them to convert the feedback into corrective measures. The author suggests that a visual representation of the problematic phonemes, using waveform and mouth/tongue position diagrams, such as the "Talk to me" software used by Hincks (2003), would make it easier for students to self-correct their errors. As animated diagrams or video instruction would be expected to offer better remedial guidance, having the ability to directly link to such media by clicking on a phoneme in the feedback would provide greater independence to the student in terms of pronunciation practice in the absence of a native speaking instructor.

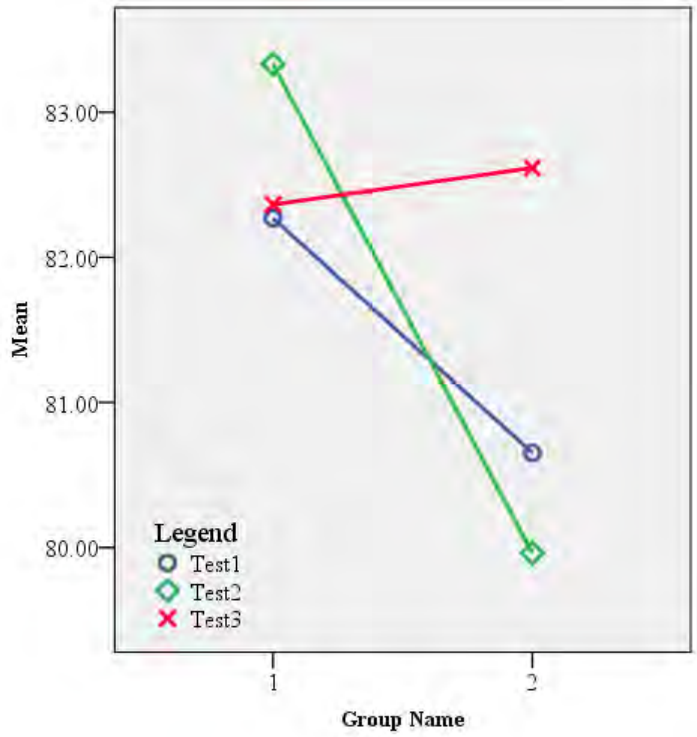
Another aspect that may have caused a weaker correlation between practise and test scores is that each practice exercise consisted of an entire phrase rather than individual words or phonemes. The practise effect may have been a more significant influence if the interface had been able to identify the problematic words and phonemes for each student and allow individuals to practise each in isolation rather than as part of a phrase.

It is noteworthy that a decrease in pronunciation accuracy was observed in each group during the period when they were not using the interface (Figure 3). This decrease could be attributed to a general lack of pronunciation practise while not using the interface, which is reflective of the findings of other studies relating to pronunciation pedagogy (Derwing et al., 2012; Derwing & Munro, 2005; Fraser, 2000; Gilakjani, 2011; Hariri, 2012).

Overall, it was found that Group 2 displayed significantly higher improvement than Group 1 between Test 1 and Test 3, (t (89) = -2.055, p = .043, 95% CI [-3.682, -0.570]). It could be argued that recency effect contributed to Group 2's better performance in Test 3 as Group 1 had not practised with the software during the six weeks prior to the third test. However, as

pronunciation scores for Group 2 in Test 2 were observed to decrease below what was considered their baseline level (Test 1), it is felt that other factors, besides practice, may have had an influence on student pronunciation in this study.

Figure 3
Mean Test Scores



Recommendations

A concerning observation regarding students' pronunciation in this study was the deterioration in their accuracy during periods when they did not have access to the practice interface. The implications for this suggest that pronunciation practice receives insufficient classroom time, or students lack environments where they can practise independently outside of the classroom. Students in this study had four hours of contact per week in the classroom with a native speaker, which significantly reduces their frequency of practice and rate at which they can develop their speaking proficiency. As suggested by Derwing et al. (2012), the practice of pronunciation should form an integral part of all EFL teaching rather than be taught independently from other skills. Based on observations and the research findings, the author echoes this suggestion and recommends that teachers adopt existing ASR applications and incorporate them into their independent study and flipped classroom sessions. Teachers should be mindful of their students' ability to interpret and respond to feedback from ASR systems to prevent them from feeling incapable of improving despite repeated attempts.

Another issue raised in this research was the quality and effectiveness of corrective feedback. ASR systems can significantly facilitate the evaluation and feedback process in terms of processing time, consistency, and protecting students' emotions. However, those systems intended for independent study purposes need to consider the user's ability to interpret feedback and make the appropriate corrections to their pronunciation. In the case of this study, the numerical feedback given by the SpeechAce interface, while adequate to identify problem

areas, does not appear to have been sufficiently user friendly to enable the students to effectively self-correct their mistakes. This shortfall can potentially demotivate the student if they fail to achieve higher pronunciation scores despite multiple attempts. Future research should look into more effective ways for independent study ASR systems to provide corrective feedback and how best to incorporate remedial practice tasks that pinpoint the user's specific pronunciation problems. A more intelligent system should detect a problematic phoneme and present the user with remedial practice specific to it. However, a future study should also investigate any significant benefit in presenting evaluation scores to the user in graphical or numerical form.

Based on the unusual decrease in pronunciation accuracy observed in Group 2 during the first round of this study, future research should investigate the factors that may lead to such reduced accuracy, which, as proposed by the author, may go beyond insufficient pronunciation practice. A future study should also examine potential gender differences in practising with the SpeechAce interface.

Conclusion

Pronunciation plays an essential part in effective communication (Fraser, 2000) and is often assessed in isolation during the standardised tests that EFL students often must complete as part of induction screening or course evaluation processes. As discussed in the research literature, teacher and student problems regarding feedback stem from limited contact time and how and when teachers should give feedback. The potential for loss of face is considered a prime reason for students not wishing to receive corrective feedback in front of their peers. It is hoped that advances in artificial intelligence software and mobile technology will create a clear pathway that offers students applications that can identify pronunciation errors and provide corrective instruction through user-friendly interfaces.

Limitations

Although the findings in this research reflect a positive outcome of using ASR for practicing pronunciation, two factors may have influenced the results, namely sample and time. Limitations on transactions permitted through the SpeechAce subscription forced a restriction in terms of the number of participants. First-year students were selected because it was felt they would have less fossilisation of errors and were also heterogeneous in their prior pronunciation training. However, this meant the sample was relatively small and predominantly female, which prevented any meaningful analysis by gender. Several studies in the research literature reported gender differences in different aspects of pronunciation accuracy (Hariri, 2012; Hincks, 2003; Jahandar et al., 2012; Khamkhien, 2010), which may also have influenced the results in this investigation. Time constraints on this study meant that students had a relatively short time using the technology and learning how to interpret the feedback scores to self-correct their pronunciation errors.

Declarations

Funding

Walailak University, Thailand, funded this work. Research grant: WU63203.

Conflicts of Interest

In my ethical obligation as a researcher, I am reporting that I am employed by and received funding from Walailak University. The University had no involvement in the design,

collection, analysis, or interpretation of this article's data or its writing. The University grants the author permission to submit this article for publication.

References

Aiello, J., & Mongibello, A. (2019). Supporting EFL learners with a virtual environment: A focus on L2 pronunciation. *Journal of e-Learning and Knowledge Society*, 15, 95–108. https://doi.org/10.20368/1971-8829/1444

Ayulistya, T. (2016). Improving students' pronunciation through Tell Me More pronunciation software. *Indonesian EFL Journal*, *2*(2), 110–116. https://doi.org/10.25134/ieflj.v2i2.643

Brett, D. (2004). Computer generated feedback on vowel production by learners of English as a second language. *ReCALL*, *16*(1), 102–113. https://doi.org/10.1017/S0958344004000813

Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge: Cambridge University Press.

Christiansen, T. (2011). Fluency and pronunciation in the assessment of grammatical accuracy in spoken production: An empirical study. *Lingue e Linguaggi, 6*, 7–20. https://doi.org/10.1285/i2239-0359v6p7

De Jong, N. H. (2018). Fluency in second language testing: Insights from different disciplines. *Language Assessment Quarterly, 15*(3), 237–254. https://doi.org/10.1080/15434303.2018.1477780

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly, 39*(3), 379–397. https://doi.org/10.2307/3588486

Derwing, T. M., Diepenbroek, L. G., & Foote, J. A. (2012). How well do general-skills ESL textbooks address pronunciation? *TESL Canada Journal, 30*(1), 22–44. https://doi.org/10.18806/tesl.v30i1.1124

Dewaele, J.-M. (2007). Multilinguals' language choice for mental calculation. *Intercultural Pragmatics, 4*(3), 343–376. https://doi.org/10.1515/IP.2007.017

Dewaele, J.-M., Furnham, A., & Petrides, K. V. (2008). Effects of trait emotional Intelligence and sociobiographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language Learning, 58*(4), 911–960. https://doi.org/10.1111/j.1467-9922.2008.00482.x

Donovan, L. A., & Macintyre, P. D. (2004). Age and sex differences in willingness to communicate, communication apprehension, and self-perceived competence. *Communication Research Reports, 21*(4), 420-427. https://doi.org/10.1080/08824090409360006

Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach, *Perspectives on fluency* (pp. 287–313). University of Michigan.

Fraser, H. (2000). Coordinating improvements in pronunciation teaching for adult learners of English as a second language. Retrieved July 12, 2020, from https://helenfraser.com.au/wp-content/uploads/ANTA-REPORT-FINAL.pdf

Gerald, K. (2000). *How to teach pronunciation* (1st ed.). Pearson Education Limited.

Gilakjani, A. P. (2011). A study on the situation of pronunciation instruction in ESL/EFL classrooms. *Journal of Studies in Education, 1*(1), 1–15.

Gilakjani, A. P. (2012). A study of factors affecting EFL learners' English pronunciation learning and the strategies for instruction. *International Journal of Humanities and Social Science, 2*(3), 119–128.

Gumbaridze, J. (2013). Error correction in EFL speaking classrooms. *Procedia - Social and Behavioral Sciences, 70*, 1660–1663. https://doi.org/10.1016/j.sbspro.2013.01.237

Haggag, H. M. (2018). Teaching phonetics using a mobile-based application in an EFL context. *European Scientific Journal, 14*(4), 189–204. https://doi.org/10.19044/esj.2018.v14n14p189

Hariri, M. (2012). A review of literature: A gender-based study of pronunciation accuracy. Research Journal of Applied Sciences, *Engineering and Technology, 4*(22), 4861–4864.

Heift, T., & Schulze, M. (2007). *Errors and intelligence in computer-assisted language learning*. New York: Routledge. https://doi.org/10.4324/9780203012215

Hincks, R. (2003). Speech technologies for pronunciation feedback and evaluation. *ReCALL*, 3–20. https://doi.org/10.1017/S0958344003000

Huang, X., & Jia, X. (2016). Corrective feedback on pronunciation: Students' and teachers' perceptions. *International Journal of English Linguistics, 6*(6), 245–254. https://doi.org/10.5539/ijel.v6n6p245

International English Language Testing System. (n.d.). IELTS Speaking band descriptors. Retrieved March 15, 2020, from https://www.ielts.org/-/media/pdfs/speaking-band-descriptors.ashx?la=en

Jahandar, S., Khodabandehlou, M., Seyedi, G., & Abadi, R. M. (2012). A gender-based approach to pronunciation accuracy of advanced EFL learners. *International Journal of Scientific & Engineering Research, 3*(6).

Kanoksilpatham, B. (2007). Navigating pathways to success in ELT. *Journal of English Studies, 3*, 6–26.

Kenworthy, J. (1987). *Teaching English pronunciation*. New York: Longman.

Khamkhien, A. (2010). Thai learners' English pronunciation competence: Lesson learned from word stress assignment. *Journal of Language Teaching and Research, 1*(6), 757–764. https://doi.org/10.4304/jltr.1.6.757-764

Kim, I. S. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching. *Journal of Educational Technology & Society, 9*(1), 322–334.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement, 30*(3), 607–610. https://doi.org/10.1177/001316447003000308

Lai, Y.-S., Tsai, H.-H., & Yu, P.-T. (2009). A multimedia English learning system using HMMs to improve phonemic awareness for English learning. *Journal of Educational Technology & Society, 12*(3), 266–281.

Latha, M. B., & Ramesh, P. (2012). Teaching English as a second language: Factors affecting learning speaking skills. *International Journal of Engineering Research & Technology, 1*(7), 1º6.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387–417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x

Leong, L., & Ahmadi, S. M. (2017). An analysis of factors influencing learners' English speaking skill. *International Journal of Research in English Education, 2*(1), 34–41. https://doi.org/10.18869/acadpub.ijree.2.1.34

Lord, G. (2005). (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania, 88*(3), 557–567. https://doi.org/10.2307/20063159

Ngamkaiwan, C. (2018). Encouraging ESL learners to speak English in large classes towards the combining arrangement activities. *Journal of Advances in Humanities and Social Sciences, 4*(3), 137–153. https://doi.org/10.13140/RG.2.2.30804.45444/1

Olson, D. J. (2014). Phonetics and technology in the classroom: A practical approach to using speech analysis software in second-language pronunciation instruction. *Hispania, 97*(1), 47–68. https://doi.org/10.1353/hpn.2014.0030

Richards, J., Platt, J., & Weber, H. (1985). *Longman dictionary of applied linguistics*. London: Longman.

Salkind, N. J. (2012). *Exploring research* (8th ed.). Pearson.

Setter, J., & Jenkins, J. (2005). State-of-the-art review article: Pronunciation. *Language Teaching, 38*(1), 1–17. https://doi.org/10.1017/S026144480500251X

Srikanth, R., Bo, L., & Salsman, J. (2012). Automatic pronunciation evaluation and mispronunciation detection using CMUSphinx. Proceedings of the *Workshop on speech and language processing tools in education* (pp. 61–68). Coling.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16(*3), 371–391. doi:10.1093/applin/16.3.371

Tavakoli, P., & Hunter, A. (2018). Is fluency being neglected in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research, 22*(3), 330–349. https://doi.org/10.1177/1362168817708462

Terrell, T. D. (1997). A natural approach to second language acquisition and learning. *Modern Language Journal, 61*(7), 325-337. https://doi.org/10.2307/324551

Yang, Y. (2014a). Is the speaking fluency strand necessary for the college students to develop in the EFL class? *Theory and Practice in Language Studies, 4*(2), 225-231. https://doi.org/10.4304/tpls.4.2.225-231

Yang, Y. (2014b). The development of speaking fluency: The 4/3/2 technique for the EFL learners in China. *International Journal of Research Studies in Language Learning, 3*(4), 55–70. https://doi.org/10.5861/ijrsll.2013.624

Corresponding author: Simon Moxon
Email: simon@sitamoxon.com